Leveraging User Annotations in Sentiment Summarization

Ryan McDonald

Joint work with Ivan Titov (Geneva), Sasha Blair-Goldensohn, Kerry Hannan, Tyler Neylon, Jeff Reynar (Google)
User Generated (Text) Content
User Generated (Text) Content
User Generated (Text) Content

Abraham Lincoln

Cafe D’Alsace

Good restaurant on UES

Cafe D’Alsace is one of the best restaurants I have found on the upper east side. The food is very good, even the simple things like a hangar steak and fries is cooked well and tasty. The cheese appetizer is excellent, though I wish they would occasionally change some of the options. Great beer and wine selection, bears mostly from bottle. Though they say you should make reservations, if there is just two of you, they will almost certainly squeeze you in without too much of a wait. Any more and I would definitely make sure you reserve.

All reviews

Cafe D’Alsace - May 24, 2006
Will Upper East Siders get their fill of choucroute gari and baeckoffe in the neighborhood formerly home to a bevy of German restaurants? Strasbourg meets Schaller und Weber? …

Was this review helpful? Yes - No
More from NewYorkCity.com »

Cafe D’Alsace Restaurant New York New...
Upper East Siders, starved for well-prepared food in a sleek setting, had often migrated downtown when their stomachs rumbled. But Cafe D’Alsace gives locals one less reason to …

Was this review helpful? Yes - No
More from Gayot.com »

natural language processing blog

to NLP machine_learning blog ... saved by
Google Scholar edit / delete
google research reference ... saved by
Picasa Web Albums edit / delete
googlenews by corel other people ...
Google Page Creator edit / delete
googlenews by corel other people ...
Google Groups edit / delete
googlenews by corel other people ...
Google Docs & Spreadsheets edit / delete
googlenews by corel other people ...
Google Account edit / delete
googlenews by corel other people ...
Home Page edit / delete
googlenews by corel other people ...

« earlier | later » showing all 8 items
User Generated (Text) Content
User Generated (Text) Content
User Generated (Text) Content
User Generated Content

• Abundant source of information
• Diverse source (expert to novice to biased to spam)
• Data often contains structured labels
What to do with annotations?

- Traditional view: use them as training data
  - Train a model and run/eval it on new data
  - Isn’t this just a contrived task?
- Not always:
  - Train sentiment classifiers on reviews use it 4 blogs
  - Train review ratings, apply on phrases or sentences
  - Train on one blog, apply to unannotated blogs
Leverage Annotations for Related Problems

- Can we use ... ?
  - Star ratings to predict phrase level sentiment
  - **Star ratings to segment the text**
  - Del.icio.us tags to place ads / improve ranking
  - Helpfulness rankings to extract QA pairs
  - Helpfulness rankings to build language models
  - ...

- Auxiliary tasks are closely related to signals provided by the user
Text Segmentation

- Focus on models for segmenting text
- Can we use aspect ratings?

<table>
<thead>
<tr>
<th>Food: 5; Decor: 5; Service: 5; Value: 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>The chicken was great. On top of that our service was excellent and the price was right. Can’t wait to go back!</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Food: 2; Decor: 1; Service: 3; Value: 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>We went there for our anniversary. My soup was cold and expensive plus it felt like they hadn’t painted since 1980.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Food: 3; Decor: 5; Service: 4; Value: 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>The food is only mediocre, but well worth the cost. Wait staff was friendly. Lot’s of fun decorations.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Food</th>
<th>“The chicken was great”, “My soup was cold”, “The food is only mediocre”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decor</td>
<td>“it felt like they hadn’t painted since 1980”, “Lots of fun decorations”</td>
</tr>
<tr>
<td>Service</td>
<td>“service was excellent”, “Wait staff was friendly”</td>
</tr>
<tr>
<td>Value</td>
<td>“the price was right”, “My soup was cold and expensive”, “well worth the cost”</td>
</tr>
</tbody>
</table>
Sentiment Summarization

- Take a set of reviews for an entity and summarize them
- **Aspect-based summarization** (Hu & Liu 2004)
  - Summarize along key aspects

**Nikos’ Fine Dining**

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Rating</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>4/5</td>
<td>“Best fish in the city”, “Excellent appetizers”</td>
</tr>
<tr>
<td>Decor</td>
<td>3/5</td>
<td>“Cozy with an old world feel”, “Too dark”</td>
</tr>
<tr>
<td>Service</td>
<td>1/5</td>
<td>“Our waitress was rude”, “Awful service”</td>
</tr>
<tr>
<td>Value</td>
<td>5/5</td>
<td>“Good Greek food for the $”, “Great price!”</td>
</tr>
</tbody>
</table>

- Many real world manual examples, e.g., Zagat.com
• Hu and Liu ’04
  • Aspect-based summarization
  • String-based aspects + lexicon sentiment
• Popescu and Etzioni ’05: Opine system
• Gammon et al. ’05
  • Aspect clusters: use most frequent word label
• Carenini ’06
  • String-based + ontologies
• Mei et al. ’07
  • Generative topic-sentiment models (at document level)
### Nikos’ Fine Dining

<table>
<thead>
<tr>
<th>Category</th>
<th>Rating</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>4/5</td>
<td>“Best fish in the city”, “Excellent appetizers”</td>
</tr>
<tr>
<td>Decor</td>
<td>3/5</td>
<td>“Cozy with an old world feel”, “Too dark”</td>
</tr>
<tr>
<td>Service</td>
<td>1/5</td>
<td>“Our waitress was rude”, “Awful service”</td>
</tr>
<tr>
<td>Value</td>
<td>5/5</td>
<td>“Good Greek food for the $$”, “Great price!”</td>
</tr>
</tbody>
</table>
Three Tasks

• Identify Aspects
  • Often we know this (pros-cons, tech specs, ontologies)

<table>
<thead>
<tr>
<th>Nikos’ Fine Dining</th>
<th>Score</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>4/5</td>
<td>“Best fish in the city”, “Excellent appetizers”</td>
</tr>
<tr>
<td>Decor</td>
<td>3/5</td>
<td>“Cozy with an old world feel”, “Too dark”</td>
</tr>
<tr>
<td>Service</td>
<td>1/5</td>
<td>“Our waitress was rude”, “Awful service”</td>
</tr>
<tr>
<td>Value</td>
<td>5/5</td>
<td>“Good Greek food for the $”, “Great price!”</td>
</tr>
</tbody>
</table>
Three Tasks

- Identify Aspects
  - Often we know this (pros-cons, tech specs, ontologies)
- Extract Mentions
  - We always have to do this

### Nikos’ Fine Dining

<table>
<thead>
<tr>
<th>Category</th>
<th>Score</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>4/5</td>
<td>“Best fish in the city”, “Excellent appetizers”</td>
</tr>
<tr>
<td>Decor</td>
<td>3/5</td>
<td>“Cozy with an old world feel”, “Too dark”</td>
</tr>
<tr>
<td>Service</td>
<td>1/5</td>
<td>“Our waitress was rude”, “Awful service”</td>
</tr>
<tr>
<td>Value</td>
<td>5/5</td>
<td>“Good Greek food for the $”, “Great price!”</td>
</tr>
</tbody>
</table>
Three Tasks

- Identify Aspects
  - Often we know this (pros-cons, tech specs, ontologies)
- Extract Mentions
  - We always have to do this
- Aggregate Sentiment
  - Again, we often know this (star ratings, eg, TripAdvisor)

Nikos’ Fine Dining

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Rating</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>4/5</td>
<td>“Best fish in the city”, “Excellent appetizers”</td>
</tr>
<tr>
<td>Decor</td>
<td>3/5</td>
<td>“Cozy with an old world feel”, “Too dark”</td>
</tr>
<tr>
<td>Service</td>
<td>1/5</td>
<td>“Our waitress was rude”, “Awful service”</td>
</tr>
<tr>
<td>Value</td>
<td>5/5</td>
<td>“Good Greek food for the $”, “Great price!”</td>
</tr>
</tbody>
</table>
Three Tasks

- Identify Aspects
  - Often we know this (pros-cons, tech specs, ontologies)
- Extract Mentions
  - We always have to do this
- Aggregate Sentiment
  - Again, we often know this (star ratings, eg, TripAdvisor)

### Nikos’ Fine Dining

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Rating</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>4/5</td>
<td>“Best fish in the city”, “Excellent appetizers”</td>
</tr>
<tr>
<td>Decor</td>
<td>3/5</td>
<td>“Cozy with an old world feel”, “Too dark”</td>
</tr>
<tr>
<td>Service</td>
<td>1/5</td>
<td>“Our waitress was rude”, “Awful service”</td>
</tr>
<tr>
<td>Value</td>
<td>5/5</td>
<td>“Good Greek food for the $”, “Great price!”</td>
</tr>
</tbody>
</table>
Aspect Identification and Extraction

- Common method: **String-based extraction**
- Find frequently occurring nouns that are modified by opinion words
- Take top K as relevant aspects
- Extract all sentences / phrases that match
- **Problem**: Get a long list of aspects w/ no clustering

**Nikos’ Fine Dining**

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Rating</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>4/5</td>
<td>“Best fish in the city”, “Excellent appetizers”</td>
</tr>
<tr>
<td>Decor</td>
<td>3/5</td>
<td>“Cozy with an old world feel”, “Too dark”</td>
</tr>
<tr>
<td>Service</td>
<td>1/5</td>
<td>“Our waitress was rude”, “Awful service”</td>
</tr>
<tr>
<td>Value</td>
<td>5/5</td>
<td>“Good Greek food for the $$”, “Great price!”</td>
</tr>
</tbody>
</table>
Aspect Identification and Extraction

- String-based example: restaurants
- Is list really summarization?
- How far down to get “cozy”, “fish”, “$”, “waitress”, “dark”?
- We really want to cluster these

Nikos’ Fine Dining

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>4/5</td>
<td>“Best fish in the city”, “Excellent appetizers”</td>
</tr>
<tr>
<td>Decor</td>
<td>3/5</td>
<td>“Cozy with an old world feel”, “Too dark”</td>
</tr>
<tr>
<td>Service</td>
<td>1/5</td>
<td>“Our waitress was rude”, “Awful service”</td>
</tr>
<tr>
<td>Value</td>
<td>5/5</td>
<td>“Good Greek food for the $”, “Great price!”</td>
</tr>
</tbody>
</table>
Topic Models

- Studied in ML and Data Mining
- LSA, PLSA, LDA, Pachinko Allocation, ...
- Build semantic “topics” of data collections
  - e.g., newsgroups into “religion”, “politics”, “science”, ...
- Simple hypothesis
  - Topics in reviews correspond to clustered aspects
- We will focus on LDA type models (Blei et al. ’03)
  - Others produce similar observations
LDA

- Generative model of text
- Sample multinomial word distributions for each topic
- The for each document $d$:

  - choose distribution of topics $\theta_d \sim \text{Dir}(\alpha)$
  - for each word $i$ in document $d$
    - choose topic $z_{d,i} \sim \theta_d$,
    - choose word $w_{d,i} \sim \varphi_{z_{d,i}}$. 
Side Note: Inference

- All methods use collapsed Gibbs (Griffiths & Steyvers ’04)
- A sample from the chain used to approx:
  - Distribution of words in topics
  - Distribution of topics in text fragments
- We tried variational techniques, but they didn’t work
- Not going to go into details
- See Titov & McDonald (WWW 2008) for more
LDA

- Problem with LDA (and most other topic models)
- Co-occurrences modeled at document level
- Topics are about instances not aspects
  - e.g., iPod versus Creative Labs
- Often clusters are meaningless

(Service??)  Topic 0: product player did support bought work unit problem $  
(Creative Labs) Topic 1: gigabeat deleted waiting jukebox creative playback  
(iPod)  Topic 11: ipod apple mac firewire dock itunes x display aac

Most topics are incoherent. Only 4 out of first 40 can be viewed as aspects.
LDA

• Simple solutions: LDA over sentences
  • Co-occurrence counts too sparse
  • Can use sliding window, but results look like LDA
  • Still can’t distinguish aspect topics from the rest

• Another solution: Multi-grain topic models
  • Model local topics (aspects) and global topics (types)
  • Creates a bottleneck for local topics
  • Words generated from sliding window
MG-LDA

- Draw global topic word dist.
- Draw local topic word dist.
- For each document d:
  - Choose a distribution of global topics $\theta_d^{gl} \sim Dir(\alpha^{gl})$.
  - For each sentence $s$ choose a distribution $\psi_{d,s}(v) \sim Dir(\gamma)$.
  - For each sliding window $v$
    - choose $\theta_{d,v}^{loc} \sim Dir(\alpha^{loc})$,
    - choose $\pi_{d,v} \sim Beta(\alpha^{mix})$.
  - For each word $i$ in sentence $s$ of document $d$
    - choose window $v_{d,i} \sim \psi_{d,s}$,
    - choose $r_{d,i} \sim \pi_{d,v_{d,i}}$,
    - if $r_{d,i} = gl$ choose global topic $z_{d,i} \sim \theta_{d}^{gl}$,
    - if $r_{d,i} = loc$ choose local topic $z_{d,i} \sim \theta_{d,v_{d,i}}^{loc}$,
    - choose word $w_{d,i}$ from the word distribution $\varphi_{z_{d,i}}^{r_{d,i}}$. 
First 8 Local Topics!!

<table>
<thead>
<tr>
<th>Sound Quality</th>
<th>Features</th>
<th>PC Connection</th>
<th>Tech Problems</th>
<th>Looks</th>
<th>Controls</th>
<th>Battery</th>
<th>Accessor’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>sound quality headphones volume bass headphones ear rock settings</td>
<td>games features clock contacts calendar alarm notes game quiz</td>
<td>usb pc windows port transfer computer mac software cable</td>
<td>reset noise backlit slow freeze turn remove playing hot</td>
<td>case pocket silver screen plastic clip easily small blue</td>
<td>button play track menu song buttons volume album tracks</td>
<td>battery hours life batteries charge aaa rechargeable time power</td>
<td>usb cable headphones adapter remote plug power charger included</td>
</tr>
</tbody>
</table>

First 4 Global Topics

<table>
<thead>
<tr>
<th>iPod</th>
<th>Creative Zen</th>
<th>Sony Walkman</th>
<th>Video Players</th>
</tr>
</thead>
<tbody>
<tr>
<td>ipod</td>
<td>zen creative Zen</td>
<td>sony walkman</td>
<td>video</td>
</tr>
<tr>
<td>music</td>
<td>micro touch xtra pad</td>
<td>memory stick sonicstage players</td>
<td>screen</td>
</tr>
<tr>
<td>apple</td>
<td>songs use mini very just</td>
<td>touch xtra pad</td>
<td>video</td>
</tr>
<tr>
<td>songs</td>
<td>use mini very just</td>
<td>stick sonicstage players</td>
<td>screen</td>
</tr>
<tr>
<td>use</td>
<td>mini very just itunes</td>
<td>sonicstage players</td>
<td>video</td>
</tr>
<tr>
<td>mini</td>
<td>very just itunes</td>
<td>players</td>
<td>screen</td>
</tr>
<tr>
<td>very</td>
<td>just itunes</td>
<td>sticks</td>
<td>video</td>
</tr>
<tr>
<td>just</td>
<td>itunes</td>
<td>sonicstage</td>
<td>video</td>
</tr>
<tr>
<td>itunes</td>
<td></td>
<td>players</td>
<td>video</td>
</tr>
</tbody>
</table>

LDA
- 40 topics
- Only 4 aspect topics
- A couple other coherent topics
- Good topics in no order
- Mostly junk topics
First 8 Local Topics!!

<table>
<thead>
<tr>
<th>Sound Quality</th>
<th>Features</th>
<th>PC Connection</th>
<th>Tech Problems</th>
<th>Looks</th>
<th>Controls</th>
<th>Battery</th>
<th>Accessor’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>sound quality</td>
<td>games</td>
<td>usb</td>
<td>reset</td>
<td>case</td>
<td>button</td>
<td>battery</td>
<td>usb</td>
</tr>
<tr>
<td>headphones</td>
<td>features</td>
<td>pc</td>
<td>noise</td>
<td>pocket</td>
<td>play</td>
<td>hours</td>
<td>cable</td>
</tr>
<tr>
<td>volume</td>
<td>clock</td>
<td>windows</td>
<td>back light</td>
<td>silver</td>
<td>track</td>
<td>life</td>
<td>headphones</td>
</tr>
<tr>
<td>bass</td>
<td>contacts</td>
<td>port</td>
<td>slow</td>
<td>screen</td>
<td>menu</td>
<td>batteries</td>
<td></td>
</tr>
<tr>
<td>earphones</td>
<td>calendar</td>
<td>transfer</td>
<td>freeze</td>
<td>plastic</td>
<td>song</td>
<td>charge</td>
<td>adapter</td>
</tr>
<tr>
<td>ear</td>
<td>alarm</td>
<td>computer</td>
<td>turn</td>
<td>clip</td>
<td>buttons</td>
<td>aaa</td>
<td>remote</td>
</tr>
<tr>
<td>rock</td>
<td>notes</td>
<td>mac</td>
<td>remove</td>
<td>easily</td>
<td>volume</td>
<td>rechargeable</td>
<td>plug</td>
</tr>
<tr>
<td>settings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

First 4 Global Topics

- Works for many other domains
- Works for both LDA and PLSA
- Empirically better than LDA on classification tasks
- Titov and McDonald 2008 at WWW

- 40 topics
- Only 4 aspect topics
- A couple other coherent topics
- Good topics in no order
- Mostly junk topics

• Works for many other domains
• Works for both LDA and PLSA
• Empirically better than LDA on classification tasks
• Titov and McDonald 2008 at WWW
MG-LDA Results

- Clearly the set of topics is better than standard models
- But, we don’t know topic labels a priori
- **Solution**
  - Let user annotations guide us
- Many things at our disposal
  - Tech specs
  - Pros-cons lists
  - **Aspect Ratings**
Aspect Ratings

- Available on an increasing number of websites
- Give us two things
  - Important aspects
  - Signals that are correlated to the text
Aspect Ratings

- Available on an increasing number of websites
- Give us two things
  - Important aspects
  - Signals that are correlated to the text

Idea
Correlate topics with user provided aspect rankings
Topics are labeled!!
Topic quality should be much better!!
Supervised LDA

- Can augment topic models to generate **observed signals**
- S-LDA (Blei and McAuliffe ’07 NIPS)
- Use document labels to guide topic construction
- We take this insight and extend it to MG-LDA
- For each aspect rating
  - Add a MaxEnt classifier to the model
  - Associate one topic to each classifier
  - MaxEnt classifier uses only words from that topic to predict rating
Multi-Aspect Sentiment Model

Can be any topic model
Multi-Aspect Sentiment Model

Can be any topic model

Overall sentiment variable
Models fact that aspect rankings are correlated
Multi-Aspect Sentiment Model

Can be any topic model

Overall sentiment variable
Models fact that aspect rankings are correlated

If we optimize the models jointly then topics will correspond directly to aspects
Multi-Aspect Sentiment Model

- Hotel model with three aspects
  - service, location, rooms
- Tied first three topics to these aspects ratings
- Trained on 10,000 reviews
- Topics correspond to associated aspects!!
Multi-Aspect Sentiment Model

- Semantic models of aspects ... what does this buy us?
- Can use model directly to find mentions of aspects

**Food: 5; Decor: 5; Service: 5; Value: 5**
The chicken was great. On top of that our service was excellent and the price was right. Can’t wait to go back!

**Food: 2; Decor: 1; Service: 3; Value: 2**
We went there for our anniversary. My soup was cold and expensive plus it felt like they hadn’t painted since 1980.

**Food: 3; Decor: 5; Service: 4; Value: 5**
The food is only mediocre, but well worth the cost. Wait staff was friendly. Lot’s of fun decorations.

<table>
<thead>
<tr>
<th>Food</th>
<th>“The chicken was great”, “My soup was cold”, “The food is only mediocre”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decor</td>
<td>“it felt like they hadn’t painted since 1980”, “Lots of fun decorations”</td>
</tr>
<tr>
<td>Service</td>
<td>“service was excellent”, “Wait staff was friendly”</td>
</tr>
<tr>
<td>Value</td>
<td>“the price was right”, “My soup was cold and expensive”, “well worth the cost”</td>
</tr>
</tbody>
</table>
Aspect Mention Extraction

Compared to a supervised MaxEnt model

**Location**

**Service**

- **Precision** vs **Recall**
  - Topic model
  - Max-ent classifier
Multiple Topics per Aspect Classifier

- Required when an aspect is diverse
- e.g., Rooms = bed, bathroom, noise, view, ...

<table>
<thead>
<tr>
<th>Rooms</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>rooms</td>
<td>room</td>
<td>room</td>
<td>room</td>
</tr>
<tr>
<td>clean</td>
<td>noise</td>
<td>clean</td>
<td>floor</td>
</tr>
<tr>
<td>hotel</td>
<td>night</td>
<td>bed</td>
<td>view</td>
</tr>
<tr>
<td>room</td>
<td>street</td>
<td>comfortable</td>
<td>rooms</td>
</tr>
<tr>
<td>small</td>
<td>did</td>
<td>rooms</td>
<td>suite</td>
</tr>
<tr>
<td>nice</td>
<td>air</td>
<td>bathroom</td>
<td>got</td>
</tr>
<tr>
<td>comfortable</td>
<td>rooms</td>
<td>small</td>
<td>views</td>
</tr>
<tr>
<td>modern</td>
<td>door</td>
<td>beds</td>
<td>given</td>
</tr>
<tr>
<td>good</td>
<td>open</td>
<td>nice</td>
<td>quite</td>
</tr>
</tbody>
</table>

- check | room  | bathroom
- arrived | noise | room
- time | night | shower
- rooms | day | night | tv
- comfortable | bed | airport |
- small | early | did | small
- beds | room | air | water
- nice | luggage | rooms | towels
- bathroom | took | noisy | bath
Aspect Mention Extraction

Multiple topics per aspect

Rooms

- 1 topic
- 2 topics
- 3 topics
- 4 topics
- log. regression
Topic Models & Sentiment Summ.

• Can topic models be used to summarize sentiment?
  • Yes!! For aspect identification and mention extraction
  • MG-LDA accurately finds aspect-like topics
    • But suffers from cluster labeling problem
  • Can augment MG-LDA to leverage aspect ratings
    • Ratings present in many data sets
    • Correlates topics w/ known aspects
    • Improves quality of topics
    • Yields highly precise mention extractors
Three Tasks

- **Identify Aspects**
  - Often we know this (pros-cons lists, star ratings)
- **Extract Mentions**
  - We always have to do this
- **Aggregate Sentiment**
  - We often know this (star ratings, eg, TripAdvisor)
  - But there is still a lot of data w/o this

---

**Nikos’ Fine Dining**

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Rating</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>4/5</td>
<td>“Best fish in the city”, “Excellent appetizers”</td>
</tr>
<tr>
<td>Decor</td>
<td>3/5</td>
<td>“Cozy with an old world feel”, “Too dark”</td>
</tr>
<tr>
<td>Service</td>
<td>1/5</td>
<td>“Our waitress was rude”, “Awful service”</td>
</tr>
<tr>
<td>Value</td>
<td>5/5</td>
<td>“Good Greek food for the $”, “Great price!”</td>
</tr>
</tbody>
</table>
Aggregate Sentiment

• Simple
  • Extract mentions for each aspect
  • Average sentiment over each of them

• Problem
  • Current sentiment classifiers are either:
    • Domain specific
    • Low in accuracy
Domain Independent Classifiers

- Build weighted semantic graph -- $A = (a_{ij})$ -- from WordNet
- Use synonyms, antonyms
- Like Hu and Liu ’04 and Kim and Hovy ’06, but with optimization
- Use label propagation from seed sets of positive, negative and neutral words
Domain Independent Classifiers

- Build weighted semantic graph \( A = (a_{ij}) \) from WordNet
- Use synonyms, antonyms
- Like Hu and Liu ’04 and Kim and Hovy ’06, but with optimization
- Use label propagation from seed sets of positive, negative and neutral words

\[
\begin{align*}
\mathbf{s}^0_i = & \begin{cases} 
+1 & \text{if } w_i \in P \\
-1 & \text{if } w_i \in N \\
0 & \forall w_i \in \text{WordNet} - P \cup N
\end{cases}
\end{align*}
\]
Domain Independent Classifiers

- Build weighted semantic graph -- $A = (a_{ij})$ -- from WordNet
- Use synonyms, antonyms
- Like Hu and Liu '04 and Kim and Hovy '06, but with optimization
- Use label propagation from seed sets of positive, negative and neutral words

$$s_i^0 = \begin{cases} 
+1 & \text{if } w_i \in P \\
-1 & \text{if } w_i \in N \\
0 & \forall w_i \in \text{WordNet}
\end{cases}$$

$$a_{ij} = \begin{cases} 
1 + \lambda & \text{if } i == j \\
+\lambda & \text{if } w_i \in \text{syn}(w_j) \land w_i \not\in M \\
-\lambda & \text{if } w_i \in \text{ant}(w_j) \land w_i \not\in M \\
0 & \text{otherwise.}
\end{cases}$$
Domain Independent Classifiers

- Build weighted semantic graph \( A = (a_{ij}) \) from WordNet
- Use synonyms, antonyms
- Like Hu and Liu ’04 and Kim and Hovy ’06, but with optimization
- Use label propagation from seed sets of positive, negative and neutral words

\[
\begin{aligned}
  s^0_i &= \begin{cases} 
    +1 & \text{if } w_i \in P \\
    -1 & \text{if } w_i \in N \\
    0 & \forall w_i \in \text{WordNet} - (P \cup N)
  \end{cases} \\
  a_{ij} &= \begin{cases} 
    1 + \lambda & \text{if } i = j \\
    +\lambda & \text{if } w_j \in \text{syn}(w_i) \& w_i \not\in M \\
    -\lambda & \text{if } w_i \in \text{ant}(w_j) \& w_j \not\in M \\
    0 & \text{otherwise}
  \end{cases}
\end{aligned}
\]

\[
\text{for } m := 1 \text{ to } M \\
  s^m := A \cdot s^{m-1}
\]
Domain Independent Classifiers

- Build weighted semantic graph -- $A = (a_{ij})$ -- from WordNet
- Use synonyms, antonyms
- Like Hu and Liu ’04 and Kim and Hovy ’06, but with optimization
- Use label propagation from seed sets of positive, negative and neutral words

\[
s^0_i = \begin{cases} 
+1 & \text{if } w_i \in P \\
-1 & \text{if } w_i \in N \\
0 & \forall w_i \in \text{WordNet} - (P \cup N)
\end{cases}
\]

\[
a_{ij} = \begin{cases} 
1 + \lambda & \text{if } i == j \\
+\lambda & \text{if } w_i \in \text{syn}(w_j) \& w_i \not\in M \\
-\lambda & \text{if } w_i \in \text{ant}(w_j) \& w_i \not\in M \\
0 & \text{otherwise.}
\end{cases}
\]

for $m := 1$ to $M$

\[
s^m = A^{m-1} \cdot s^0
\]

raw-score($x$) := $\sum_{i=1}^{n} s_i$. 
Meta Classifier

• Collect scores for:
  • Sentence / phrase
  • Previous & next sentence / phrase
  • Document

• Train a classifier on a labeled set of sentences/phrases
  • Use scores as features
Meta Classifier

• Collect scores for:
  • Sentence / phrase
  • Previous & next sentence / phrase
  • Document

• Train a classifier on a labeled set of sentences/phrases

• Use scores as features

![Bar graph showing positive and negative scores for WordNet, Meta, and Meta + User. Positive scores are around 62.9 and negative scores are around 54.7.]
Meta Classifier

- Collect scores for:
  - Sentence / phrase
  - Previous & next sentence / phrase
  - Document

- Train a classifier on a labeled set of sentences/phrases

- Use scores as features

![Bar chart showing positive and negative scores for WordNet, Meta, Meta + User categories. Positive scores are 62.9, 68.6, and 68.5, respectively. Negative scores are 54.7 for all categories.](image-url)
• Collect scores for:
  • Sentence / phrase
  • Previous & next sentence / phrase
  • Document
• Train a classifier on a labeled set of sentences/phrases
• Use scores as features

User reviews usually have **overall** sentiment

Overall sentiment highly correlated w/ phrase & sentence sentiment
Meta Classifier

- Collect scores for:
  - Sentence / phrase
  - Previous & next sentence / phrase
  - Document
- Train a classifier on a labeled set of sentences/phrases
- Use scores as features

Add feature for overall user rating

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>WordNet</td>
<td>62.9</td>
<td>54.7</td>
</tr>
<tr>
<td>Meta</td>
<td>68.6</td>
<td>68.5</td>
</tr>
<tr>
<td>Meta + User</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Meta Classifier**

- Collect scores for:
  - Sentence / phrase
  - Previous & next sentence / phrase
  - Document
- Train a classifier on a labeled set of sentences/phrases
- Use scores as features

Add feature for overall user rating
Summary

• Aspect-based sentiment summarization

• Can use topic models and standard sent classifiers

• Quality improves by leveraging correlated user signals
  • Aspect ratings for aspect mention extraction
  • Overall ratings for phrase/sentence classification

• Models generalize to any segmentation problem where there are correlated signals
  • e.g., del.icio.us bookmarks, blog labels, helpfulness, ...
Thanks

- Joint work with Ivan Titov, Sasha Blair-Goldensohn, Kerry Hannan, Tyler Neylon, George Reis and Jeff Reynar
- Thank you to Kenji and Tsujii lab for invite